

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 17, 2003, 17:30:02 ; Search time 591.3 seconds
(without alignments)
1383.718 Million cell updates/sec

Title: US-09-827-383B-3
Perfect score: 20
Sequence: 1 caccgacgcttaataagtaag 20

Scoring table:- IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues 5777422

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	16.8	84.0	10822	1	AE004517	AE004517 Pseudomon
C 2	16.8	84.0	76730	8	AP004533	AP004533 Lotus jap
C 3	16.8	84.0	161261	9	CNS05TE3	AL358332 Human chr
C 4	16.8	84.0	220760	2	AC097210	AC097210 Rattus no
C 5	16.8	84.0	231901	2	AC123417	AC123417 Rattus no
C 6	16.8	84.0	240009	2	AC096200	AC096200 Rattus no
C 7	16.4	82.0	113164	2	AC136925	AC136925 Oryza sat
C 8	16.4	82.0	177887	2	AC134256	AC134256 Oryza sat
C 9	16.4	82.0	217841	2	AC095995	AC095995 Rattus no
C 10	16.4	82.0	241206	2	AC120484	AC120484 Rattus no
C 11	15.8	79.0	802	11	BV032188	BV032188 S212P6348
C 12	15.8	79.0	41125	10	AC002108	AC002108 Genomic s
C 13	15.8	79.0	71516	9	AF117830	AF117830 Homo sapi
C 14	15.8	79.0	89350	8	ATT12H17	AL021635 Arabidops
C 15	15.8	79.0	108612	10	AL844490	AL844490 Mouse DNA
C 16	15.8	79.0	146640	2	AC114335	AC114335 Canis fam
C 17	15.8	79.0	161053	9	AL356608	AL356608 Human DNA
C 18	15.8	79.0	169865	9	AC079967	AC079967 Homo sapi
C 19	15.8	79.0	172071	9	AC012676	AC012676 Homo sapi
C 20	15.8	79.0	177726	9	AC004083	AC004083 Homo sapi
C 21	15.8	79.0	181835	9	AC091152	AC091152 Homo sapi
C 22	15.8	79.0	186608	10	AL844178	AL844178 Mouse DNA
C 23	15.8	79.0	189043	8	AP004366	AP004366 Oryza sat
C 24	15.8	79.0	195007	2	AC137037	AC137037 Rattus no
C 25	15.8	79.0	197775	9	AC012366	AC012366 Homo sapi
C 26	15.8	79.0	199577	8	ATCHRIV57	AL161557 Arabidops
C 27	15.8	79.0	201386	2	AC135720	AC135720 Mus muscu
C 28	15.8	79.0	207304	2	AC107146	AC107146 Rattus no
C 29	15.8	79.0	210688	10	AC127371	AC127371 Mus muscu
C 30	15.8	79.0	214556	2	AC128105	AC128105 Rattus no
C 31	15.8	79.0	218077	2	AC097219	AC097219 Rattus no
C 32	15.8	79.0	234379	2	AC109379	AC109379 Rattus no
C 33	15.8	79.0	238106	2	AC123207	AC123207 Rattus no
C 34	15.8	79.0	238106	2	AC123207	AC123207 Rattus no
C 35	15.8	79.0	239945	2	AC094050	AC094050 Rattus no
C 36	15.8	79.0	248254	2	AC133448	AC133448 Rattus no
C 37	15.8	79.0	255270	2	AC130767	AC130767 Rattus no
C 38	15.4	77.0	12375	1	AF425233	AF425233 Yersinia
C 39	15.4	77.0	74925	8	AP006434	AP006434 Lotus jap
C 40	15.4	77.0	132092	2	AC135422	AC135422 Oryza sat
C 41	15.4	77.0	233851	2	AC097870	AC097870 Rattus no
C 42	15.4	77.0	270707	2	BX323558	BX323558 Danio rer
C 43	15.2	76.0	298	6	AR250541	AR250541 Sequence
C 44	15.2	76.0	507	8	AF222834	AF222834 Phaeoisar
C 45	15.2	76.0	659	9	HSA338788	AJ338788 Homo sapi

ALIGNMENTS

RESULT 1
AE004517/c
LOCUS
DEFINITION
Pseudomonas aeruginosa PAO1, section 78 of the complete genome.
ACCESSION
AE004517 AE004091
VERSION
AE004517.1 GI:9946710
KEYWORDS
SOURCE
ORGANISM
Pseudomonas aeruginosa PAO1
Pseudomonas aeruginosa PAO1
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 10822)
AUTHORS
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.G., Kowalik,D.J.,

AE004517 10822 bp DNA linear BCT 19-FEB-2003
Pseudomonas aeruginosa PAO1, section 78 of the complete genome.

Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., and Paulsen, I.T.
Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen

Nature 406 (6799), 959-964 (2000)

20437337

10984043

2 (bases 1 to 10822)

Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University Of Washington, Box 352145, Seattle, WA 98195, USA

3 (bases 1 to 10822)

Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)

Direct Submission

Submitted (04-FEB-2003) Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia V5A 1S6, Canada

This represents the February 3, 2003 version of the continually updated, reviewed, *Pseudomonas aeruginosa* PA01 genome annotation, from PseudoCAP (see <http://www.pseudomonas.com> for latest updates and links to alternate annotations). PseudoCAP is coordinated by Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert E.W. Hancock (University of British Columbia, Canada). We welcome submission through www.pseudomonas.com of any proposed changes.

'Protein name confidence' is used to rate our confidence of the accuracy of the protein name.
Class 1: Function experimentally demonstrated in *P. aeruginosa*.
Class 2: Function of highly similar gene experimentally demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known).
Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene.
Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.

Location/Qualifiers

1. 10822
/organism="Pseudomonas aeruginosa PA01"

/mol_type="genomic DNA"

/strain="PA01"

/db_xref="taxon:208964"

81..506

/locus_tag="PA0817"

81..506

/locus_tag="PA0817"

/note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)"

/codon_start=1

/transl_table=11

/product="probable ring-cleaving dioxygenase"

/protein_id="AAG04206.1"

/db_xref="GI:9946711"

/translation="MSLMPFHLAIPVHDLPAARFYEYFVGLSEGRSAEHWDFDFEG
HQLVIHQPTDSQRHAGSNPVDHVPVPHFVGLAWDDWHAERLQQRGTRFVIE
PYIRFKQVGEQATLFLFDPCGNALFEKSFDRMGQLFAK"

complement (525..797)

/locus_tag="PA0818"

CDS

complement (525..797)
/locus_tag="PA0818"
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAG04207.1"

/db_xref="GI:9946712"

/translation="MPYRVLLILALTPALASSATLDERIGDLEHAFKPDLDGTTSAK
PASPPPERFKLAPLGSVVRKKECATSSLPDEQSSCALEFCTDCAQ"

953..1246

/locus_tag="PA0819"

953..1246

/locus_tag="PA0819"

/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAG04208.1"

/db_xref="GI:9946713"

/translation="MAPCGKTVTAQSPSTFIPLRRTSQCNQHPGCATLYLDCSAASE
DLFDTARTLESAIGTSLVDLLDETEEPGESTCRITGSLLSLFEAYARRHR"

1545..2357

/locus_tag="PA0820"

1545..2357

/locus_tag="PA0820"

/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAG04209.1"

/db_xref="GI:9946714"

/translation="MPAVIAENDVSIWDDGTGAVHYPRYSAILKPGVQVYTKGL
TKSFADQRLSAQHYFGIARIGKVDPDESSKGLDIPALIESFPFEQAVLAKEGNSY
LEPIPPKASNYWRSVREIGQVTFDSILARAEIKPQVAESWGDALGFESRTEGS
LSTYGVRYERDPTLRQAIHIGVACKACGDFGKAYGEYAKGFIHVHTQPISEFE
APKAVNPETDLVPLCANHAVVHRRDRRTLSVDELKGLVGRWVCAEADVVV"

complement (2464..3552)

/locus_tag="PA0821"

complement (2464..3552)

/locus_tag="PA0821"

/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAG04210.1"

/db_xref="GI:9946715"

/translation="MEQFFDLAAELAKQPHLEIAGNLLMKSGPEDIYIGVLCRGFTL
YFKAHTPLVRESLCQCFDEPERLAEPHLTWLWEEPAQCKPLTAYRDTQPLREMWGA
MDEDDHLSFCYTSKGKSRDAGAWLFDIYGRSQAKMGHDLVLEFSVPLTYQERQPL
DFLQLFIDFARRLEPEQYAGHAYNLSPSWNDPESEAFMAARMPGLDVGTACLLAN
TPEFKPTRIKTVSWLTNNRLALAGGLDALRAQPSHSHFAYRYGDDGVVTOAGAYP
YIAGDAEDSRPAPYVLLNHALKGIRYETIGSLHCGSHDGLRLVGVMAADQWLKRLDVE
DSEIPRWCDKLLSAEYLDATNTLPERL"

complement (3569..4015)

/locus_tag="PA0822"

complement (3569..4015)

/locus_tag="PA0822"

/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAG04211.1"

/db_xref="GI:9946716"
/translation="MTKNPPQILDTSQGNPHGWIPGWISKYWDDEPHPPFKPGKG
MIRRPDVIIVQNPRTQDNIKQVEMKFPDPDPHNREQLDYAALAGNKNKIVEMKP
SDCDCQENQRKVPVEQAGWAVAIAGGVNVLTRGRSPRPMIPAY"
complement(4037..4270)
/locus_tag="PA0823"
complement(4037..4270)
/locus_tag="PA0823"
/note="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04212.1"
/db_xref="GI:9946717"
/translation="MSDQPSLSRGGMSPEGRTPVGTLEPRLDPPQKKVLCSAVCHCS
STPNLSQDGSLKQACVAQRLGELDEVLQGRSP"
complement(4267..4785)
/locus_tag="PA0824"
complement(4267..4785)
/locus_tag="PA0824"
/note="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04213.1"
/db_xref="GI:9946718"
/translation="MSGKPAARVTDPTTCVPVGHGSPNIVQSPDVVFDGLPAARQGD
ASACGSPMISAVSSTVLINGLPVAVTLGSAHGNWVGSGTVLIGDVFTPAPRAPAL
PLNRNSVPCSGRFQLIDHETGKPVAGRRVRVWSSGGWNAFDTTDDGWTSWIERPTAE
ILYIDLVRQCD"
complement(4959..5294)
/locus_tag="PA0825"
complement(4959..5294)
/locus_tag="PA0825"
/note="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
Query Match 84.0%; Score 16.8; DB 1; Length 10822;
Best Local Similarity 90.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CACCGACGCTAATAGTTAAG 20
Db 1355 CACAGACGCTCATAGTTAAG 1336
RESULT 2
AP004533 76730 bp DNA linear PLN 14-DEC-2001
LOCUS
DEFINITION
Lotus japonicus genomic DNA, chromosome 3, clone:LjTl4G02, TM0080,
complete sequence.
ACCESSION
AP004533
VERSION
AP004533.1 GI:17736900
KEYWORDS
HTG.
SOURCE
Lotus japonicus
ORGANISM
Lotus japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosoids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
Lotus.
REFERENCE
1
AUTHORS
Sato,S., Kaneko,T., Nakamura,Y., Asamizu,E., Kato,T. and Tabata,S.
TITLE
Structural Analysis of a Lotus japonicus Genome. I. Sequence
Features and Mapping of Fifty-six TAC clones which cover the 5.4 Mb
Regions of the Genome
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 76730)
AUTHORS
Nakamura,Y.

TITLE
JOURNAL
Direct Submission
Submitted (13-DEC-2001) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
URL:http://www.kazusa.or.jp, Tel:81-438-52-3935,
Fax:81-438-52-3934)
FEATURES
Location/Qualifiers
source
1..76730
/organism="Lotus japonicus"
/mol_type="genomic DNA"
/db_xref="taxon:34305"
/chromosome="3"
/clone="LjTl4G02"
/clone_lib="LjT library"
/note="TAC clone:TM0080"
BASE COUNT 25039 a 13819 c 14263 g 23609 t
ORIGIN
Query Match 84.0%; Score 16.8; DB 8; Length 76730;
Best Local Similarity 90.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CACCGACGCTAATAGTTAAG 20
Db 16636 CACCGACGCTAATATTTAG 16655
RESULT 3
CNS05TE3/c 161261 bp DNA linear PRI 13-JUL-2001
LOCUS
DEFINITION
Human chromosome 14 DNA sequence BAC R-255G12 of library RPCI-11
from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION
AL358332
VERSION
AL358332.3 GI:14787750
KEYWORDS
HTG; HTGS ACTIVEFIN.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 161261)
AUTHORS
Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
Gyapay,G., Saurin,W. and Weissensbach,J.
Sequencing of the human chromosome 14
Unpublished
REFERENCE
2 (bases 1 to 161261)
AUTHORS
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (13-JUL-2001) Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On Jul 17, 2001 this sequence version replaced gi:10432564.
COMMENT
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-1047D15
Downstream BAC (overlapping the SP6 end) : R-102G14 (AC=AL122125)
----- Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 8.58x in Q20 bases; sum-of-contigs

Overall quality chart :
Range : bases
0 :
1 - 9 : 2
10 - 19 : 6
20 - 29 : 57

30 - 39 : 189
40 - 49 : 2710
50 - 59 : 8979
60 - 69 : 11203
70 - 79 : 22101
80 - 89 : 44981
90 - 99 : 71033

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES

source

Location/Qualifiers

1. 161261
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-255G12"
/clone_lib="RPCI-11"
126666..126895
/note="matching EMBL:252148
RHdb:RH42652
RHdb:RH31310
dbSTS:STS2223
Identified using the e-PCR software (G. Schuler)"
154125..154250
/note="matching EMBL:T86731
RHdb:RH53680
RHdb:RH32887
dbSTS:STS26145
Identified using the e-PCR software (G. Schuler)"
BASE COUNT 45445 a 31846 c 34019 g 49951 t
ORIGIN

Query Match

Best Local Similarity 84.0%; Score 16.8; DB 9; Length 161261;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1 CACCGACGCTAATAGTTAAG 20

Db

10507 CACCAACTCTAATAGTTAAG 10488

RESULT 4

AC097210/c

LOCUS

DEFINITION

AC097210 220760 bp DNA linear HTG 10-MAY-2003
Rattus norvegicus clone CH230-207E1, WORKING DRAFT SEQUENCE, 2
unordered pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC097210.7 GI:30521257
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 220760)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allene, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpthy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuhewa, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasan, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 220760)

Worley, K. C.

Direct Submission

JOURNAL

Submitted (12-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 220760)

Rat Genome Sequencing Consortium.

Direct Submission

JOURNAL

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On May 10, 2003 this sequence version replaced gi:24942582.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GIMY

Center clone name: CH230-207E1

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 209257 bases at least Q40

Consensus quality: 211357 bases at least Q30

Consensus quality: 212520 bases at least Q20

Estimated insert size: 216663; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

```

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 219368: contig of 219368 bp in length
* 219369 219468: gap of unknown length
* 219469 220760: contig of 1292 bp in length.
FEATURES             source
    source
    1..220760
        /organism="Rattus norvegicus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10116"
        /clone="CH230-207E1"
    1..1836
        /note="wgs_end_extension"
        clone_end:Sp6"
    5729..6382
        /note="clone boundary"
        clone_end:Sp6
    site:EcoRI
    end_sequence:BH289343"
    complement(216545..217346)
        /note="clone boundary"
        clone_end:T7
    site:EcoRI
    end_sequence:RBWKI25TJB"
BASE COUNT 61737 a 43269 c 43471 g 64832 t 7451 others
ORIGIN
Query Match      84.0%;   Score 16.8;   DB 2;   Length 220760;
Best Local Similarity 90.0%;   Pred. NO. 1.7e+02;
Matches 18;   Conservative 0;   Mismatches 2;   Indels 0;   Gaps 0;

QY 1 CACCGACGCTAATAGTTAAG 20
    ||||| ||||| ||||| |||||
Db 175855 CACCTACTCTAATAGTTAAG 175836

RESULT 5
AC123417
LOCUS AC123417 231901 bp DNA linear HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-195F22, WORKING DRAFT SEQUENCE.
ACCESSION AC123417
VERSION AC123417.4 GI:25088183
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 231901)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,

```

```

Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C.,
Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 231901)
Worley,K.C.
Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 231901)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23265439.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXDR
Center clone name: CH230-195F22
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 212992 bases at least Q40
Consensus quality: 215592 bases at least Q30

```

Consensus quality: 216777 bases at least Q20
Estimated insert size: 216147; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 231901: contig of 231901 bp in length.

FEATURES

source

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

BASE COUNT 58062 a 48279 c 49404 g 61493 t 14663 others
ORIGIN

Query Match 84.0%; Score 16.8; DB 2; Length 231901;
Best Local Similarity 90.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CACCGACGCTAATAGTTAAG 20

Db 130601 CACACACGCTAATAGTTAAG 130620

RESULT 6

AC096200

LOCUS

DEFINITION AC096200 Rattus norvegicus clone CH230-37K11, WORKING DRAFT SEQUENCE, 3
unordered pieces.

ACCESSION

VERSION AC096200.6 GI:30522427

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 240009)

AUTHORS

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,

Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louisedge,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,J., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 240009)

Worley,K.C.

Direct Submission

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 240009)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On May 10, 2003 this sequence version replaced gi:24818305.
The sequence in this assembly is a combination of BAC-based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

Center project name: GEOL
Center clone name: CH230-37K11
----- Summary Statistics

Assembly program: Atlas 3.0;
Consensus quality: 226448 bases at least Q40
Consensus quality: 228877 bases at least Q30
Consensus quality: 230092 bases at least Q20
Estimated insert size: 233615; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 237591: contig of 237591 bp in length
* 237592 237691: gap of unknown length
* 237692 238737: contig of 1046 bp in length
* 238738 238837: gap of unknown length
* 238838 240009: contig of 1172 bp in length.

FEATURES

source

1. .240009
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-37K11"

misc_feature

1. .1403
/note="wgs end_extension
clone_end:T7"

misc_feature

5429. .7521
/note="wgs end_extension
clone_end:T7"

misc_feature

7557. .8401
/note="clone_boundary
clone_end:T7"

misc_feature

site:EcoRI
end_sequence:BH305843"
141219. .225407
/note="clone_boundary
clone_end:Sp6"

BASE COUNT

62638 a 49403 c 51000 g 68223 t 8745 others

ORIGIN

Query Match 84.0%; Score 16.8; DB 2; Length 240009;
Best Local Similarity 90.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CACCGACGCTAATAGTTAAG 20

Db 220049 CACACACGCTAATAGTTAAG 220068

RESULT 7

AC136925/c

LOCUS

DEFINITION Oryza sativa (japonica cultivar-group) chromosome 11 clone
OSJNBa0090B19, *** SEQUENCING IN PROGRESS ***; 7 ordered pieces.

ACCESSION

AC136925

VERSION

KEYWORDS

SOURCE

ORGANISM

AC136925.1 GI:24850539

HTG; HTGS PHASE2.

Oryza sativa (japonica cultivar-group)

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE

AUTHORS

1 (bases 1 to 113164)
Pal,A.K., Dixit,A., Sureshababu,K., Singh,A., Pal,S., Yadav,M.,
Gaikwad,K., Ghazi,I.A., Swain,S.C., Srivastava,S., Bhargava,A.,
Dalal,V., Batra,K., Sharma,T.R., Mohapatra,T. and Singh,N.K.
Genomic sequence for Oryza sativa chromosome 11

TITLE

JOURNAL

Unpublished

REFERENCE

AUTHORS

2 (bases 1 to 113164)
Pal,A.K., Dixit,A., Sureshababu,K., Singh,A., Pal,S., Yadav,M.,
Gaikwad,K., Ghazi,I.A., Swain,S.C., Srivastava,S., Bhargava,A.,
Dalal,V., Batra,K., Sharma,T.R., Mohapatra,T. and Singh,N.K.
Direct Submission

TITLE

JOURNAL

Submitted (11-NOV-2002) IRCS, NRC on Plant Biotechnology, Indian
Agricultural Research Institute, LBS Centre, New Delhi, Delhi
110012, India

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 1160: contig of 1160 bp in length
* 1161 1260: gap of unknown length
* 1261 37630: contig of 36370 bp in length
* 37631 37730: gap of unknown length
* 37731 50155: contig of 12425 bp in length
* 50156 50255: gap of unknown length
* 50256 52672: contig of 2417 bp in length
* 52673 52772: gap of unknown length
* 52773 64423: contig of 11651 bp in length
* 64424 64523: gap of unknown length
* 64524 69586: contig of 5063 bp in length
* 69587 113164: contig of 43478 bp in length.

FEATURES

source

Location/Qualifiers
1. .113164
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39947"
/chromosome="11"
/clone="OSJNBa0090B19"
/note="(japonica cultivar-group)"

BASE COUNT 34095 a 22866 c 22420 g 33175 t 608 others

ORIGIN

Query Match 82.0%; Score 16.4; DB 2; Length 113164;
Best Local Similarity 94.4%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCGACGCTAATAGTTAA 19

Db 40250 ACCGACGCTAATAGTTAA 40233

RESULT 8

AC134256

LOCUS

DEFINITION Oryza sativa (japonica cultivar-group) chromosome 11 clone
OSJNBa0033H01, *** SEQUENCING IN PROGRESS ***; 7 ordered pieces.

ACCESSION

AC134256

VERSION

AC134256.3 GI:28604227

KEYWORDS

HTG; HTGS PHASE2.

SOURCE

Oryza sativa (japonica cultivar-group)

ORGANISM

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

REFERENCE AUTHORS

1 (bases 1 to 177887)
Bhargava, A., Pal, S., Batra, K., Yadav, M., Sureshbabu, K., Singh, A., Srivastava, S., Swain, S.C., Dixit, A., Ghazi, I.A., Dalal, V., Pal, A.K., Gaikwad, K., Sharma, T.R., Mohapatra, T. and Singh, N.K.
Genomic sequence for Oryza sativa chromosome 11

TITLE JOURNAL

Unpublished
2 (bases 1 to 177887)
Bhargava, A., Pal, S., Batra, K., Yadav, M., Sureshbabu, K., Singh, A., Srivastava, S., Swain, S.C., Dixit, A., Ghazi, I.A., Dalal, V., Pal, A.K., Gaikwad, K., Sharma, T.R., Mohapatra, T. and Singh, N.K.
Direct Submission
Submitted (24-SEP-2002) IIRGS, NRC on Plant Biotechnology, Indian Agricultural Research Institute, LBS Centre, New Delhi, Delhi 110012, India

TITLE JOURNAL

REFERENCE AUTHORS

3 (bases 1 to 177887)
Bhargava, A., Pal, S., Batra, K., Yadav, M., Sureshbabu, K., Singh, A., Srivastava, S., Swain, S.C., Dixit, A., Ghazi, I.A., Dalal, V., Pal, A.K., Gaikwad, K., Sharma, T.R., Mohapatra, T. and Singh, N.K.
Direct Submission
Submitted (28-FEB-2003) IIRGS, NRC on Plant Biotechnology, Indian Agricultural Research Institute, LBS Centre, New Delhi, Delhi 110012, India

TITLE JOURNAL

On Feb 28, 2003 this sequence version replaced gi:27502446.
* NOTE: This is a 'working draft' sequence. It currently

COMMENT

* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 10831: contig of 10831 bp in length
* 10832 10931: gap of unknown length
* 10932 49727: contig of 38796 bp in length
* 49728 49827: gap of unknown length
* 49828 62445: contig of 12618 bp in length
* 62446 62545: gap of unknown length
* 62546 105748: contig of 43203 bp in length
* 105749 105848: gap of unknown length
* 105849 117930: contig of 12082 bp in length
* 117931 118030: gap of unknown length
* 118031 119573: contig of 1543 bp in length
* 119574 119674: gap of unknown length
* 119674 177887: contig of 58214 bp in length.

FEATURES source

Location/Qualifiers
1..177887
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39947"
/chromosome="11"
/clone="OSJNBa0033H01"
/note="(japonica cultivar-group)"
BASE COUNT 52016 a 35110 c 36445 g 53713 t 603 others

BASE COUNT ORIGIN

Query Match 82.0%; Score 16.4; DB 2; Length 177887;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCGACGCTAATAGTTAA 19

Db 75066 ACCGACGCTAATACTTAA 75083

RESULT 9

AC095995/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-32A21, WORKING DRAFT SEQUENCE, 2
unordered pieces.

ACCESSION VERSION KEYWORDS SOURCE ORGANISM

AC095995
AC095995.7 GI:30522852
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE AUTHORS

1 (bases 1 to 217841)
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpthy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackeleme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Puozzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE JOURNAL

Unpublished
2 (bases 1 to 217841)
Worley, K.C.

REFERENCE AUTHORS

Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

TITLE JOURNAL

3 (bases 1 to 217841)
Rat Genome Sequencing Consortium.

REFERENCE AUTHORS

Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On May 10, 2003 this sequence version replaced gi:24818121.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GECQ
Center clone name: CH230-32A21
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 209555 bases at least Q40
Consensus quality: 211996 bases at least Q30
Consensus quality: 213711 bases at least Q20
Estimated insert size: 219326; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 216619: contig of 216619 bp in length
* 216620 216719: gap of unknown length
* 216720 217841: contig of 1122 bp in length.

FEATURES

source
Location/Qualifiers
1. .217841
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-32A21"
misc_feature
1. .1835
/note="wgs end_extension
clone end:T7"
3321..130068
/note="clone_boundary
clone end:T7"
site:ECORI
end_sequence:BH325857"
complement(215092..215837)
/note="clone_boundary
clone end:Sp6
site:ECORI
end_sequence:BH325858"
BASE COUNT 62710 a 46664 c 45725 g 59600 t 3142 others
ORIGIN

Query Match 82.0%; Score 16.4; DB 2; Length 217841;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCGACGCTAATAGTTAAG 20
|||||
Db 37547 CCGACTCTAATAGTTAAG 37530

RESULT 10

AC120484/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC120484 241206 bp DNA linear HTG 12-OCT-2002
Rattus norvegicus clone CH230-135H12, WORKING DRAFT SEQUENCE.
AC120484
AC120484.3 GI:23907778
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 241206)

REFERENCE

AUTHORS

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokeme,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 241206)

Worley,K.C.

Direct Submission

Submitted (07-MAY-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 241206)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (12-OCT-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Oct 12, 2002 this sequence version replaced gi:21908466.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXSF
Center clone name: CH230-135H12
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 228650 bases at least Q40
Consensus quality: 230295 bases at least Q30
Consensus quality: 231252 bases at least Q20
Estimated insert size: 235536; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* *
1 241206: contig of 241206 bp in length.

FEATURES
source

1..241206
Location/Qualifiers
1..241206
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-135H12"
820..1685
/note="clone_boundary"
clone_end:T7
site:EcoRI

misc_feature

BASE COUNT 62790 a 53676 c 52951 g 62493 t 9296 others
ORIGIN

Query Match 82.0%; Score 16.4; DB 2; Length 241206;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCGACGCTAATAGTTAAG 20
|||||

Db 127576 CCGACTCTAATAGTTAAG 127559
|||||

RESULT 11

BV032188

LOCUS

DEFINITION

S212P6348FA9.T0 CZECHII/Ei Mus musculus STS genomic, sequence tagged site.

ACCESSION

VERSION

BV032188

802 bp DNA linear STS 30-MAY-2003

SV032188

GI:31116083

KEYWORDS

SOURCE
ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

STS.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 802)

Wade,C.M., Kulbokas,E.J. III, Kirby,A.W., Zody,M.C., Mullikin,J.C., Lander,E.S., Lindblad-Toh,K. and Daly,M.J.

The mosaic structure of variation in the laboratory mouse genome Nature 420 (6915), 574-578 (2002)

22354684

12466852

Contact: Kerstin Lindblad-Toh

Whitehead Institute for Biomedical Research, Center for Genome Research

320 Charles Street, Cambridge, MA 02141, USA

Tel: 6172521477

Fax: 6172580903

Email: kersli@genome.wi.mit.edu

Primer A: None

Primer B: None

STS size: 802

Protocol:

WGS-discovery: Paired-end low-coverage whole genome shotgun reads were generated from 129S1/SvimJ, C3H/HeJ, and BALB/cByJ. The WGS reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP detection was carried out by SSAHA-SNP. 225,000 reads were annotated

as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J and the strain from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%.

FEATURES

source

1..802
Location/Qualifiers
1..802
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="CZECHII/Ei"
/db_xref="taxon:10090"
/map="- 11 22-762 26487420-26486680"
/clone_lib="CZECHII/Ei"
<1..>802

STS

BASE COUNT 208 a 156 c 217 g 221 t

ORIGIN

Query Match 79.0%; Score 15.8; DB 11; Length 802;

Best Local Similarity 89.5%; Pred. No. 7.1e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ACCGACGCTAATAGTTAAG 20
|||||

Db 52 ACAGACGCCAATAGTTAAG 70
|||||

RESULT 12

AC002108

LOCUS

DEFINITION

Genomic sequence from Mouse 4, complete sequence.

AC002108

VERSION AC002108.1 GI:2133882

KEYWORDS

SOURCE

ORGANISM

linear ROD 29-MAY-1997

41125 bp DNA

Genomic sequence from Mouse 4, complete sequence.

AC002108

AC002108.1 GI:2133882

HTG.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 41125)

Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,

Fasman,K.H. and Lander,E.S.

Genomic sequence from Mouse 4

Unpublished

2 (bases 1 to 41125)

Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,

Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,

Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Frapp,W.J.,

Gage, D., Geraigery, K., Hagos, B., Jacotot, L., Lane, M., MacKenzie, J., Marquis, N., McDermott, J., McInerney, N., Morrow, J., Nachman, A., Naylor, J., O'Connor, T., Peterson, K., Rollins, G., Spencer, J., Stilwell, J., Stone, C., Strickland, C., Sydney, K., Wilmer, F. and Zody, M.

TITLE
 Direct Submission
JOURNAL
 Submitted (13-MAY-1997) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
 3 (bases 1 to 41125)
AUTHORS
 Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W.,
 Fasman, K.H., Lander, E.S., McKernan, K., Munro, C., Richardson, P.,
 Barna, N., Chang, A., Cooke, P., Daly, M.J., Forrest, C., Fripp, W.J.,
 Gage, D., Geraigery, K., Hagos, B., Jacotot, L., Lane, M., MacKenzie, J.,
 Marquis, N., McDermott, J., Moloney, N., Morrow, J., Nachman, A.,
 Naylor, J., O'Connor, T., Peterson, K., Rollins, G., Spencer, J.,
 Stilwell, J., Stone, C., Strickland, C., Sydney, K., Wilner, F. and
 Zody, M.

2007/11/11

TITLE	Direct Submission
JOURNAL	Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On May 29, 1997 this sequence version replaced gi:2085725. The Staden databases, finishing information, and all chromatographic files used in the assembly of this clone are available from our anonymous ftp site.

All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

```

FEATURES
  source
    Location/Qualifiers
      1..41125
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /chromosome="4"
        /clone="Pla2s"
        /clone_lib="unknown"
        complement(6..528)
        /rpt_family="ORR1A-INT"
        complement(533..822)
        /rpt_family="ORR1A2"
        complement(1491..1581)
        /rpt_family="PB1D7"
        complement(2637..2990)
        /rpt_family="WTE"
        complement(3021..3154)
        /rpt_family="B1-F"
        complement(3301..3507)
        /rpt_family="MER65_internal"
        3665..3706
  repeat_region
  repeat_region
  repeat_region
  repeat_region
  repeat_region
  repeat_region
  repeat_region

```


... of clone 296C9⁺

[illegible]

```

FEATURES
  source
    location
      1. 71516
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /db_xref="taxon:9606"
        /chromosome="8"
        /map="8q21.3"

```

```

repeat_region /rpt_family="L2"
repeat_region complement(30393. .30585)
repeat_region /rpt_family="MIR"
repeat_region 30797. .32782
repeat_region /rpt_family="L2"
repeat_region 32937. .33070
repeat_region /rpt_family="L2"
repeat_region complement(35789. .35953)
repeat_region /rpt_family="L2"
repeat_region complement(36340. .36703)
repeat_region /rpt_family="THE1C"
repeat_region 37049. .37878
repeat_region /rpt_family="L2"
repeat_region 38812. .38834
repeat_region /rpt_family="POLY_A"
repeat_region 39226. .39374
repeat_region /rpt_family="MIR"
repeat_region 39375. .39559
repeat_region /rpt_family="MER45B"
repeat_region 39600. .39698
repeat_region /rpt_family="MER45C"
repeat_region 39758. .40022
repeat_region /rpt_family="MER45C"
repeat_region 40751. .41255
repeat_region /rpt_family="L2"
repeat_region complement(42153. .42241)
repeat_region /rpt_family="L2"
repeat_region 42260. .42399
repeat_region /rpt_family="MER3"
repeat_region 42747. .42799
repeat_region /rpt_family="MIR"
repeat_region 44670. .45617
repeat_region /rpt_family="L1MB8"
repeat_region 47452. .47479
repeat_region /rpt_family="POLY_A"
repeat_region 48024. .48134
repeat_region /rpt_family="MER5A"
repeat_region 48218. .48380
repeat_region /rpt_family="L2"
repeat_region 50555. .51031
repeat_region /rpt_family="LTR22"
repeat_region complement(51301. .51541)
repeat_region /rpt_family="MIR"
repeat_region 51765. .52131
repeat_region /rpt_family="THE1C"
repeat_region 52272. .52772
repeat_region /rpt_family="MLT1D"
repeat_region complement(55159. .55184)
repeat_region /rpt_family="(CA)n"
repeat_region complement(59095. .59237)
repeat_region /rpt_family="(TGA)n"
repeat_region 59261. .59434
repeat_region /rpt_family="MIR"
repeat_region 62113. .62575
repeat_region /rpt_family="L2"
repeat_region complement(65342. .65488)
repeat_region /rpt_family="MIR"
repeat_region 65602. .65900
repeat_region /rpt_family="L2"
repeat_region 65925. .66258
repeat_region /rpt_family="MLT1A1"
repeat_region 66304. .66346
repeat_region /rpt_family="(TAAA)n"
repeat_region 67132. .67177
repeat_region /rpt_family="(CAAAC)n"
repeat_region 67178. .67305
repeat_region /rpt_family="FLAM_A"
repeat_region 67306. .67319
repeat_region /rpt_family="(CAAAC)n"
repeat_region 67388. .67713
repeat_region /rpt_family="MLT1B"
repeat_region 67738. .68071
repeat_region /rpt_family="AluJo"

```

```

repeat_region complement(68073. .68278)
repeat_region /rpt_family="MIR"
repeat_region complement(68668. .68963)
repeat_region /rpt_family="AluSq"
repeat_region 69033. .69229
repeat_region /rpt_family="L2"
repeat_region 69488. .69795
repeat_region /rpt_family="AluSx"
repeat_region complement(70513. .70782)
repeat_region /rpt_family="L1ME"

BASE COUNT 22610 a 13954 c 13537 g 21415 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 9; Length 71516;
Best Local Similarity 89.5%; Pred. No. 6.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CACCGACGCTAATAGTTAA 19
Db 41897 CACAGACGCTAAGAGTTAA 41879

RESULT 14
ATT12H17 89350 bp DNA linear PLN 03-FEB-1998
LOCUS Arabidopsis thaliana DNA chromosome 4, BAC clone T12H17 (ESSAI
DEFINITION project).
ACCESSION AL021635
VERSION AL021635.1 GI:2827538
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A.,
Duesterhoeft,A., Bancroft,I., Mewes,H.W., Mayer,K. and Schueller,C.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 89350)
AUTHORS EU Arabidopsis sequencing,project.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1998) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
FEATURES
source
1. .89350
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
669. .1151
/gene="T12H17.10"
669. .1151
/gene="T12H17.10"
/note="similarity to auxin-induced protein 10A, Glycine
max., PIR2:JQ1099"
/codon_start=1
/product="predicted protein"
/protein_id="CAA16547.1"
/db_xref="GI:2827539"
/db_xref="SPTREMBL:O49643"
/translation="MKRLGFKIGHRFVKIFKWLIRSRNQTGKQCLTGILNPVTKI
YSLARRCLRRGANLRCGNKPGQTRLGNEPKTPTVPRGHLVHVHVGESGEDTRVVVPV
IYFNHPLFGELLEQAERYVGFQPGRIMIPCRVSDFEKVQVRIAWDHCRCSTFKIL
"
669. .1151
/gene="T12H17.10"
/number=1
2406. .2983
gene
exon
gene

```

CDS

/gene="T12H17.20"
join(2406. .2754,2844. .2983)
/gene="T12H17.20"
/note="similarity to predicted proteins, Arabidopsis thaliana"
/codon_start=1
/product="predicted protein"
/protein_id="CAA16548.1"
/db_xref="GI:2827540"
/db_xref="SPTREMBL:O49644"
/translation="MAYTNKVAVAVGAAVFLAVMNPRTWTEAQTYPKLDRLCVMMIPDILEECFTHDRLKPTEDCCNDLKNATMTQVDCDCLDNFLESLSFSDLRSTFSAGVLKKCDVSHKYMCOAAKNRGEAKGGRNSTTICDNSITNTSVGGKNKVATSM SAFGLVAILLFV MF"
2406. .2754
/gene="T12H17.20"
/number=1
2755. .2843
/gene="T12H17.20"
/number=1
2844. .2983
/gene="T12H17.20"
/number=2
3327. .3400
/product="tRNA-Asn"
/note="codon recognized: AAC"
4696. .5489
/gene="T12H17.30"
join(4696. .5035,5407. .5489)
/gene="T12H17.30"
/note="contains EST gb:N38564"
/codon_start=1
/product="putative protein"
/protein_id="CAA16549.1"
/db_xref="GI:2827541"
/db_xref="SPTREMBL:O49645"
/translation="MAYTNKVTISA AVATMMFLAVTIVDAQSMPPMPKFNPVICALADLPNIVQLCYENLDLTPEECCNDLKSSSTIQVNCLCDNFIAHPNSNGNISQARYDLVNSACGVADKFA CKGGDASGGSTNKIAASMVLLGLVASLFF"
4696. .5035
/gene="T12H17.30"
/number=1
5036. .5406
/gene="T12H17.30"
/number=1
5407. .5489
/gene="T12H17.30"
/number=2
6595. .7105
/gene="T12H17.40"
join(6595. .6940,7035. .7105)
/gene="T12H17.40"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAA16550.1"
/db_xref="GI:2827542"
/db_xref="SPTREMBL:O49646"
/translation="MAVAVTAVLEFLAVVIA PQWTETKKPPRPSTSDTSGTSGRDRRTMCPLSIPGIVQNCYATLNAFPSKECKDLKTASKREVTCLCNNVIAHPDPLYTNTNQV NKACGVLDKYACDVAPTFTIVTYNYDMIIAVIQDP"
6595. .6940
/gene="T12H17.40"
/number=1
6941. .7034
/gene="T12H17.40"
/number=1
7035. .7105
/gene="T12H17.40"
/number=2
8019. .9209
/gene="T12H17.50"
complement(8019. .9209)
/gene="T12H17.50"

exon

intron

exon

tRNA

gene

CDS

exon

intron

exon

gene

CDS

exon

intron

exon

gene

CDS

exon

intron

exon

gene

gene

CDS

complement(8019. .9209)
/gene="T12H17.50"
/note="similarity to predicted protein, Arabidopsis thaliana"
/codon_start=1
/product="predicted protein"
/protein_id="CAA16551.1"
/db_xref="GI:2827543"
/db_xref="SPTREMBL:O49647"
/translation="MEKNQNPNTWSDPLDLLNLVFKRLSFANFRQAKSVCSWYSASKSVPKNQIPWLMLFPKDKNNKNSSCTIFFNPEDKDQLYQTQDLGVEFAKSVCLATYKSWLLMQDSKYNLYILNPFYEKIGLPAIESQQVGMVKVDQTI DDDFLTDFDDHNVKLFRGNNTVRTPVFWIDEKTEYIALWGLGYCVCVAKNGDKLWNQIPEIILDSLDVMYKDHKLYSFSYRNLFITILDFSGEIPRKAFCQFMHVYRSEWLSRQLSNSWCVAETKLVVTVTGDVLLVERMLRHSRIQSFNVYKFYSSGTFLDKVELADSLGDEAMLLDLGITVL ANEVEGLHRNTIYFSDSHDTTKDLFLFN YETREMEPLHKFD FDCSLLELSARWFLPSFSHT"
complement(8019. .9209)
/gene="T12H17.50"
/number=1
9990. .13709
/gene="T12H17.60"
join(9990. .10023,10239. .10535,10564. .10664,10910. .11292,11755. .11809,11898. .12053,12143. .12311,12391. .12455,12534. .12674,12769. .12885,13024. .13096,13244. .13380,13468. .13525,13612. .13709)
/gene="T12H17.60"
/note="similarity to p48, Homo sapiens, PATCHX:G904032 contains EST gb:Aa605571, H76919, H76638, R30271, Z25650"
/codon_start=1
/product="HSP associated protein like"
/protein_id="CAA16552.1"
/db_xref="GI:2827544"
/db_xref="SPTREMBL:O49648"
/translation="MRRTWLAESSVAIAPLLAEPOSTWFPENTPECATVMPDLLEKCFATGSVPTEDCCTDLKSATSTQVTCLDNYIANFAVSNITGPYSKAITTRKCGVFDK YSCDGTSKVISNKGAFMVI NGFWFCVWKEKKRKEAAVAMGVLTLGLINGLIVAH GKTRQQNKPVCKLILFQFHFGFCSIRFSANKRARSKTGDAYPPFRNEKVLQKPYSLPFVPQI LEIETR LTKTKTRLLTLRPKPTLFLFLSLSRSEADFSQKDGFEAYLGA KIPFGVEEDKDKPRSFVVEESDDMDDETEEVKPKVEEEDEEIVESDVELEGDTVE PDNDPPQKMGDSSVEVTDENREAAQEA KGKAMEALSEGNFDEAIEHLTRAITLNP TSAIMYGNRASVYIKLKPNAAIRDANAAL EIPDSAKGYKSRGMARAMLGEWAEAAKDILH LASTIDYDEEISAVLKVEPN AHKLEEHRRKYDR LKREDKKAE RDLRRRAEAAQAA YDKAKKEEQSSSRPSGGFP GGMGGMPGGMGGMPGAGGGGMPGAGGGMPGGGMP GGMDFSKILNDPELMTAFSDPEVMAALQDV MKNPANLAKHQANPKVAPVIAGMMGKFA GPQ"
9990. .10023
/gene="T12H17.60"
/number=1
10024. .10238
/gene="T12H17.60"
/number=1
10239. .10535
/gene="T12H17.60"
/number=2
10536. .10563
/gene="T12H17.60"
/number=2
10564. .10664
/gene="T12H17.60"
/number=3
10665. .10909
/gene="T12H17.60"
/number=3
10910. .11292
/gene="T12H17.60"
/number=4
11293. .11754
/gene="T12H17.60"
/number=4
11755. .11809
/gene="T12H17.60"
/number=5

exon

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon


```
intron      11810..11897
            /gene="Tl2H17.60"
            /number=5
exon        11898..12053
            /gene="Tl2H17.60"

Query Match      79.0%; Score 15.8; DB 8; Length 89350;
Best Local Similarity 89.5%; Pred. No. 6.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CACCGACGCTAATAGTTAA 19
        ||||| ||||| ||||| |||||
Db      31293 CACAGACGCTATTAGTTAA 31275

Search completed: November 17, 2003, 22:06:03
Job time : 597.3 secs
```

```
RESULT 15
AL8444490/c
LOCUS      108612 bp      DNA      linear      ROD 19-SEP-2002
DEFINITION Mouse DNA sequence from clone RP23-98P13 on chromosome X, complete
            sequence.
ACCESSION  AL8444490
VERSION    AL844490.4 GI:23304656
KEYWORDS   HTG.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 108612)
AUTHORS   Ellwood,M.
TITLE     Direct Submission
JOURNAL   Submitted (19-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Sep 23, 2002 this sequence version replaced gi:22798535.
COMMENT   ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-98P13 is from the RPCI-23 Mouse PAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

Location/Qualifiers

1..108612

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="X"

/clone="RP23-98P13"

/clone_lib="RPCI-23"

```
FEATURES
source
33150 a 21828 c 21230 g 32404 t
BASE COUNT
ORIGIN
```